Reprinted from *J. Food Sci.*
47(4): 1186-1197

Perception of Texture by Trained and Consumer Panelists

ARMAND V. CARDELLO, OWEN MALLER, JOHN G. KAPSALIS, RONALD A. SEGARS, F. MILES SAWYER, CLAIRE MURPHY, and HOWARD R. MOSKOWITZ

ABSTRACT

Four experiments were conducted to assess the relationships between judgments of the perceived texture of foods by trained and consumer panelists. In Experiment 1, no differences were observed between trained texture profile panelists and naive consumers in a similarities scaling task. In Experiments 2 and 3, good linear correlations were observed between scalar judgments of texture, although a broader perceptual range was evidenced for trained panelists. In Experiment 4, psychophysical exponents of texture were found to be larger for trained than for consumer panelists, and judgments of acceptability also differed between the two groups. It was concluded that, through experience, trained texture profile panelists develop a broader perceptual range of textures, but that regression equations can be developed to relate these data to consumer data.

INTRODUCTION

SUCCESSFUL PRODUCT DEVELOPMENT in the food industry is dependent upon the proficiency of the organization's sensory evaluation group. The decisions made by this group are, in turn, dependent upon both the propitious selection of sensory panels for testing and the proficiency of these panels. The panels may be either affective (preference or acceptance) or analytical (discriminative or descriptive).

The two major factors that distinguish most sensory panels are the nature of the tasks that they perform and the nature and degree of training that they have received. Affective panels normally consist of naive and untrained consumers of the product, who evaluate the product for its hedonic qualities or its acceptability. Analytical panels, on the other hand, consist of "experienced," "trained" or "expert" panelists, who evaluate a product for some more specific aspects of its taste, odor, texture, etc.

Although the decision to use one type of panel versus another is determined by the nature of the question to be answered and the sensory/psychophysical method to be used, on occasion, the availability of a panel will become an important factor. For example, since the cost and personnel requirements to establish a wide variety of test panels is often prohibitive, especially for smaller laboratories, the situation may arise wherein a test requires a particular type of panel that is not available at the testing facility. In most cases this situation is correctly resolved by contracting the test to another laboratory. Unfortunately, under some circumstances, the panel(s) that is (are) available will be used to address the problem. When this happens, some training or re-orientation is usually undertaken to prepare the panel for the new task. However, the degree to which this training produces a panel that is truly adequate to the task is rarely examined.

Authors Cardello, Maller, Kapsalis, and Segars are affiliated with the U.S. Army Natick Research & Development Laboratories, Natick, 01760. Author Sawyer is with the Dept. of Food Science & Nutrition, Univ. of Massachusetts, Amherst, MA. Author Murphy is with the Monell Chemical Senses Center, Univ. of Pennsylvania, Philadelphia, PA. Author Moskowitz is with Developmetrics, The Weston Group, Inc. Westport, CT.

While several reports have discussed techniques for the screening, selection and training of members for various types of test panels, e.g., Martin (1973), Bressan and Behling (1977), Cross et al. (1978), Sawyer et al. (1962), Kirkpatrick et al. (1957), Swartz and Furia (1977), Zook and Wessman (1977), Girardot et al. (1952), Wittes and Turk (1968), Scholosberg et al. (1954) Civille and Szczesniak (1973), Hall et al. (1959), Gruber and Lindberg (1966), Stone et al. (1974), and Caul (1957), relatively fewer attempts have been made to quantitatively compare responses made by one type of panel to those made by another type of panel. The most significant of the early studies in this area were those conducted by Miller et al. (1955), Kiehl and Rhodes (1956), Simone et al. (1956), Peryam and Haynes (1957), Calvin and Sather (1959), Ellis (1963), and Pangborn and Dunkley (1964), in which it was generally found that laboratory consumer panel judgments of acceptability did not correspond well with consumer ratings of acceptability in the field. It was concluded that laboratory consumer panel judgments of acceptability agree in direction, but not in magnitude, with field consumer panel ratings. More recently, Kluter (1974) has reported on a study by Nichols et al. (1972) that showed rank-order correlations ranging from −0.8 to +1.0 between laboratory and field panel ratings of acceptability, depending upon the food product. Only the recent study by Moskowitz, et al. (1979a) has examined the relationship between judgments of specific sensory attributes by a laboratory consumer panel and by a trained analytical panel. In that study, data on certain relationships between consumer and trained texture profile panel judgments for rye breads were presented.

Knowledge of the relationship between consumer panel judgments and trained or expert panel judgments is important for several reasons. First is the fact that trained and expert panels are expensive to establish and maintain, while consumer panels require little formal training and members need not have highly specialized abilities. If reliable and predictive relationships can be established between judgments (either affective or analytical) made by these two types of panels, then consumer panels could be interchanged, under suitable conditions, with trained or expert panels, in order to achieve the most cost-effective use of personnel resources. Secondly, important time can be saved in the research and development cycle by employing a single panel to generate multiple types of sensory information, thereby eliminating the need to wait for results of one type of test before proceeding with another type. Thirdly, knowledge of such relationships can assist in the interpretation of discrepancies that arise between trained or expert panel data and consumer data, and help limit the extent to which data generated by a trained or expert panel poorly reflect the sensory experiences of actual consumers of the product. Lastly, from a theoretical perspective, comparison of judgments made by untrained consumer panels and trained or expert panels would make it possible to assess the experiential factors that are most important in determining sensory judgments, so that the limits of panel interchange can be delineated.

In the study of the textural attributes of food, two types of panels are most commonly employed. These are the trained profile panel, commonly modeled after the General Foods Texture Profile Panel (Brandt et al., 1963; Szczesniak et al., 1963) and the consumer panel, used frequently for obtaining hedonic measures of texture for marketing and product optimization purposes, and also used for "consumer texture profiling" (Szczesniak and Skinner, 1973; Szczesniak et al., 1975). In several recent studies conducted in our laboratory, both types of panels have been employed in the evaluation of food products, wherein both panels were instructed to perform exactly the same sensory/psychophysical tasks. These tasks included simple similarity scaling, in which only judgments of the similarity among samples was required, hedonic scaling, and scaling of specific textural attributes, using both category scaling and magnitude estimation. Comparison of the relationships between these consumer and trained panel judgments highlight the effects of training on the judgmental process of making textural evaluations and on the feasibility of using consumer panels to generate data normally obtained from trained panels, and vice versa.

EXPERIMENT 1

IN VIEW OF THE OBVIOUS difference between trained and consumer panels in their knowledge of descriptive sensory terminology, the first study was designed to assess the relationship between consumer and trained panel judgments in a task that does not require such knowledge. For this purpose, similarities scaling was chosen, because it is a simple psychophysical judgment that can be made without knowledge of specific terminology and because neither our laboratory consumers nor trained panelists had had prior experience in making this type of judgment.

Method

Panels. The first panel was a nine-member trained texture profile panel consisting of employees of the U.S. Army Natick R&D Laboratories. All members had been trained in the General Foods' Texture Profile Method, and all had served on the panel for a minimum of 1 yr. The panel training consisted of 1 wk of intensive discussion and demonstration of the basic techniques for making textural evaluations, exposure to each of the six standard scales of texture (Szczesniak et al., 1963), exposure to food items representing each of the geometrical characteristics of texture, construction of a series of increasingly complex texture profiles for a variety of food items, and training in the rendering of operational definitions of more complex texture attributes. During the time between initial training and the conduct of the test described herein, panel members met on an average of one – two times per week and were involved in the description and scaling of the textural attributes of meats, gelatins, food bars, and various other food products. The panel had never previously evaluated fish products.

The second panel was an untrained laboratory consumer panel consisting of 10 volunteer employees. None of these panelists had ever received training in making textural or other analytical sensory judgments, although many had previously participated in consumer acceptance testing of a variety of foods in the laboratory. This panel had also never evaluated fish products, performed a similarities scaling task nor used a line scaling technique.

Test samples. Six species of fish: Haddock *(Melanogrammus aeglefinus)*, Halibut *(Hippoglossus hippoglossus)*, Blackback Flounder *(Pseudopleuronectes americanus)*, White Hake *(Urophycis tenuis)*, Mackerel *(Scomber scombrus)*, and Pollock *(Pollachius virens)* were chosen as test samples. Fish fillets were selected as the test product because they had not been tested previously by either panel. All samples were purchased fresh locally, baked in foil to an internal temperature of 160°F and maintained at a constant temperature in a steam table for a period not exceeding 10 min prior to serving. Samples were presented to subjects in approximately 2-ounce servings.

Procedure. The experiment was conducted in the standard sensory testing booths of the Food Acceptance Laboratory at NLABS. Panelists were first presented with a sample of each of the six fish to acquaint them with the range of samples to be encountered. The panelists rinsed their mouth with distilled water between samples. Panelists judged the overall similarity of each fish paired once with itself and once with each of the other species, for a total of 21 pairs of samples. Responses were made by placing a slash-mark on a 30 cm line. The end points were labelled "similar" and "dissimilar." Sample pairs were presented in random order to all panelists.

Data analysis. The line-ratings of similarity were transformed to numerical values by measuring the distance from the end-point of the scale to the line marked by the subject. These proximity measures were then used to generate a "sensory map" of the fish, using multidimensional scaling procedures.

The similarity judgments for the consumer and trained panels were analyzed using ALSCAL–4 (Young and Lewyckyj, 1979). ALSCAL–4 is a metric multidimensional scaling program that uses similarity measures between pairs of stimuli and arranges the stimuli into a multidimensional space, in which the relative distances between stimulus points reflect the relative disimilarities between judged samples. The dimensionality of the space is determined by an iterative procedure in which the goodness of fit of any n-dimensional space to the data is assessed by a "Stress" value. Usually, as n increases, the stress (error) decreases, due to reduced constraint in positioning the stimuli, until a nonsignificant decrease in stress is achieved. The program also includes the capability of analyzing individual differences among subjects.

Results

The results of the application of ALSCAL–4 to the similarities data are presented in Fig. 1. The best-fitting solution was found with three dimensions and had a calculated stress of 0.15.

The three dimensional solution in Fig. 1 is intuitively appealing, because the perceptual space is easily interpretable from data collected subsequently in our laboratory on the sensory properties of fish (Kapsalis and Maller, 1981).

Dimension 1 is clearly related to the color of the cooked fish, and is best described as a light-dark dimension. Mackerel, a very dark-fleshed fish, appears at one end of this dimension, while halibut, white-hake and haddock, all very

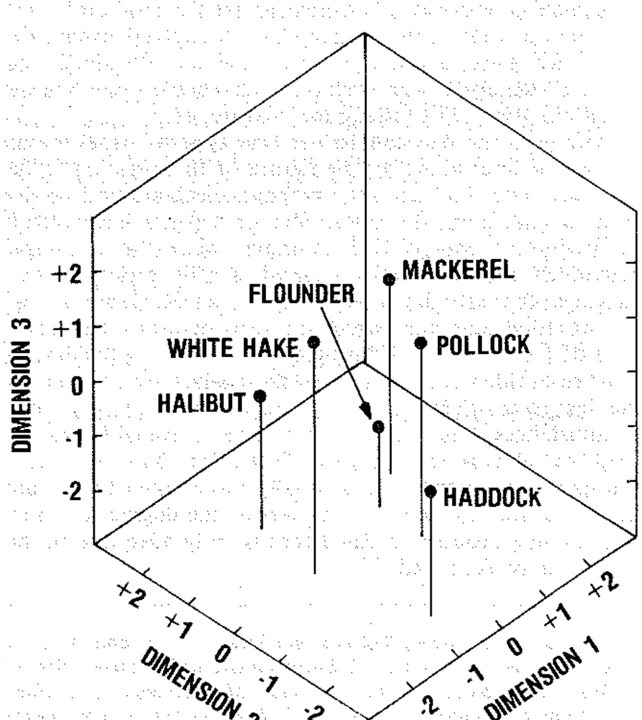

Fig. 1—Best-fitting 3-dimensional solution to the similarities data of Experiment 1. Data were analyzed by ALSCAL-4 and the 3-dimensional solution had a stress of 0.15.

white-fleshed fish, appear at the other end. Pollock and flounder, which are grayish in color, fall between these two extremes. This light-dark dimension has repeatedly been found to be one of the best discriminative variables in descriptive tests of fish conducted in our laboratory, and its emergence as a primary sensory dimension is in agreement with these data.

Dimension 2 in this space appears to be a textural dimension and is related to the perceived flakiness of the fish. Halibut, which appears as one extreme on this dimension, is a solid-muscle fish with little or no flakiness. Mackerel is also lacking in appreciable flakiness. Haddock, pollock and white hake, on the other end of this dimension, are characterized by a high degree of flakiness.

Dimension 3 appears to be related to the overall flavor intensity of the fish, with mackerel loading high on this dimension due to its high oil content and correspondingly strong fish-oil flavor, and flounder and haddock appearing at the other end, due to their mild flavor. The white hake and pollock are of surprisingly similar magnitude to the mackerel on this dimension. However, white hake has been described as possessing a characteristically strong "stale fish" or "earthy" flavor and pollock may sometimes possess an off-taste due to season of catch or feeding ground.

Fig. 2–4 show plots of the derived subject weights for each panelist on each pair of the three dimensions. These

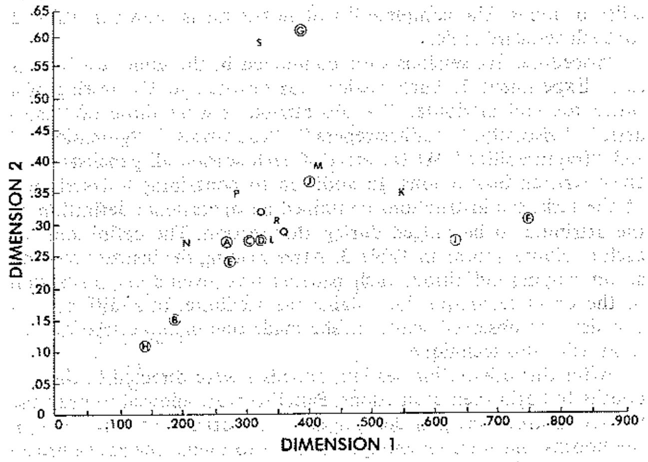

Fig. 2—Derived subject weights for trained (uncircled) and consumer (circled) panelists on Dimensions 1 and 2.

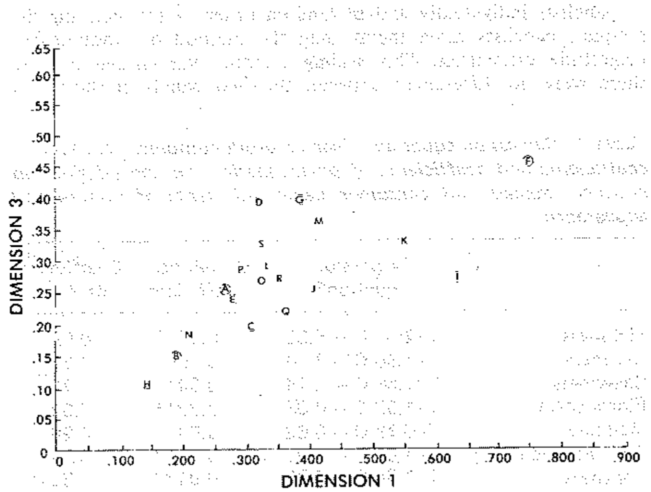

Fig. 3—Derived subject weights for trained (uncircled) and consumer (circled) panelists on Dimensions 1 and 3.

weights represent the relative degree to which each panelist used each perceptual dimension in making his judgement of overall similarity. The uncircled letters in each plot represent trained panelists, the circled letters represent consumer panelists. If either group weighted one or more dimensions differently from the other group, then the individuals within the groups would cluster in different segments of the plots. The fact that they do not indicates that the trained and consumer panelists did not differentially weight the importance of any of the three obtained dimensions. The important aspect of these data for the present discussion is that the trained texture profile panelists did not place any more weight on the textural dimension of flakiness (Dimension 2) than did the consumer panelists.

EXPERIMENT 2

THE SECOND STUDY was designed to assess the relationship between scalar judgements made by trained and consumer panelists on specific textural attributes. This task is considerably more complex than that of Experiment 1, requiring that the panelist be able to identify specific textural attributes in the product. Thus, prior experience in making such judgments may be more likely to affect panelist ratings.

Method

Panels. The first panel was a trained texture profile panel, similar to that used in Experiment 1, and comprised of six members. Although Experiment 2 was conducted approximately 18 months after Experiment 1, five of the six members were the same as in Experiment 1. In the 18 month interim, all panelists acquired extensive experience in judging the textural attributes of a wide variety of fish.

The second panel was a consumer panel similar to that described in Experiment 1, but was composed of a different random sample of 40 panelists for each test session. The selection of a different sample of consumer for each test session was made to minimize the experience that these panelists would acquire in judging the textural attributes of the test samples over the course of the experiment. All of the consumer panelists had prior experience in using category scales, but none had any known experience in judging specific textural attributes of fish or other seafood products.

Test samples. Eighteen species of fish: Whiting *(Merluccius bilinearis)*, Mackerel *(Scomber scombrus)*, White Hake *(Urophycis tenuis)*, Cusk *(Brosme brosme)*, Monkfish *(Lophius americanus)*, Pollock *(Pollachius virens)*, Tilefish *(Lopholatilus Chamealeonticeps)*, Wolffish *(Anarchichas lupus)*, Striped Bass *(Morone saxatilis)*, Blackback Flounder *(Pseudopleuronectes americanus)*, Weakfish *(Cynoscion regalsi)*, Grouper *(Mycertoperca microlepis)*, Haddock *(Melanogrammus aeglefinus)*, Halibut *(Hippoglussus hippoglossus)*,

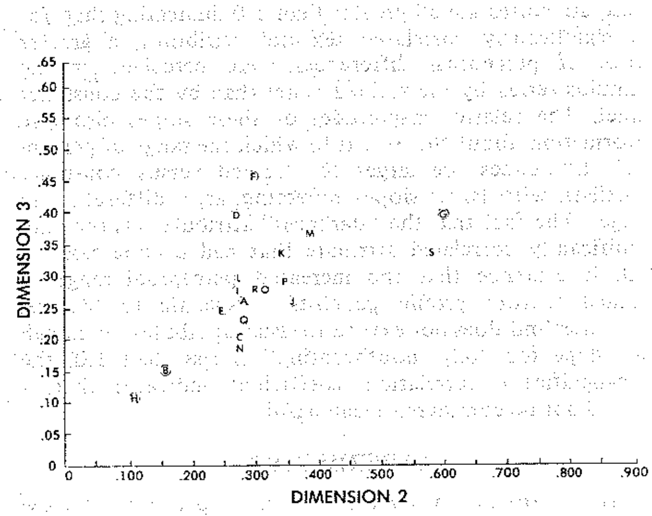

Fig. 4—Derived subjects weights for trained (uncircled) and consumer (circled) panelists on Dimensions 2 and 3.

Swordfish *(Xiphias gladius)*, Cod *(Gadus morhua)* – Scrod and market size, and Bluefish *(Pomatomus saltatrix)* served as test samples. All samples were purchased fresh locally as fillets, cut, sealed in individual plastic cooking pouches, and cooked in a retort to an internal temperature of 160°F. Samples were presented to panelists in approximately 2-ounce servings.

Procedure. Consumer tests were conducted in the same test booths as in Experiment 1 and all tests involved the presentation of a single fish sample. At the start of each session consumer panelists were given a printed response sheet with written instructions for the test. Panelists were asked to rate six separate textural attributes of the fish. These were the "hardness," "flakiness," "chewiness," "fibrousness," "moistness," and "oily mouthcoating" of the fish. In addition, each rated the visual attribute of "darkness" and six flavor attributes of the fish. Only the textural and visual attributes are discussed here, since flavor attributes were not evaluated by the texture profile panel. All attributes were rated on a 7-point category scale, with end-points labeled "slight" (1) and "strong" (7). A "none" category was also provided in the event that a panelist did not perceive any of a given attribute in the sample. Neither definitions of the attributes nor demonstration of specific techniques for judging the attributes were provided to the consumer panelists.

Trained panelists received the same samples as did the consumer panelists, rated them on the same textural and visual attributes and used the same 7-point scale. Neither group of panelists were informed about the species of fish being tested, and all panelists' judgments were made independently of one another.

Results

The mean ratings for each attribute and sample were calculated for each panel. In order to assess the relationship between judgments of the two panels, mean ratings for the trained panel were regressed against mean ratings for the consumer panel for each attribute. The slope of the obtained regression equation reflects the rate of growth of perceived magnitude for the trained panel as a function of the perceived magnitude for consumer panelists. Thus, a slope less than 1.0 indicates that for each unit increase in perceived magnitude by consumers, a smaller increase was perceived by trained panelists. A slope equal to 1.0 reflects equivalent sensory increases, and a slope greater than 1.0 indicates a more rapid growth of perceived magnitude for trained panelists.

Table 1 contains the obtained linear regression equations relating trained panel judgments to consumer panel judgments for each attribute, as well as the corresponding correlation coefficients and coefficients of determination for each.

Although the correlation coefficients are statistically significant for six of the seven attributes, indicating a significant degree of linearity between judgments made by the two panels, the slopes (b) of the regression equations for these attributes are all greater than 1.0, indicating that, for all significantly correlated textural attributes, a greater range of perceptual differences were perceived in the stimulus series by the trained panel than by the consumer panel. The relative magnitudes of these slopes also give information about the extent to which the range of perceptual differences are larger for trained versus consumer panelists, with larger slopes reflecting larger differences in ranges. The fact that the "darkness" attribute was the only significantly correlated attribute that had a slope close to 1.0, is evidence that the increased perceptual range of trained texture profile panelists is specific to textural attributes and does not extend to visual attributes. Although the slope for "oily mouthcoating" is less than 1.0, the non-significant correlation coefficient indicates that it should not be considered meaningful.

EXPERIMENT 3

WHILE THE DATA OBTAINED in Experiment 2 show differences in the sensory ratings assigned by trained and consumer judges for fish samples, it is not clear whether these effects on perceptual range are due to a general heightened awareness and/or sensitivity to these textural attributes or whether the effects are (1) specific to fish, since the texture profile panel had been evaluating fish samples for 18 months prior to the start of this experiment, (2) specific to the use of a 7-point category scaling procedure, which had also been used extensively by the trained panel, but not by consumers, or (3) simply due to differences between consumers and trained panelists in their definitions of, or techniques for evaluating, these attributes. To determine whether these factors contributed to the obtained results, the following experiment was conducted.

Method

Panels. The first panel was a trained texture profile panel similar to that used in Experiments 1 and 2. Eight members served in these tests. Five members participated in Experiments 1 and 2, three did not. The illness of one member during one series of sessions, reduced the panel size to seven on these occasions. All panelists had previous experience in using the method of magnitude estimation.

The second panel was a consumer panel similar to that described in Experiments 1 and 2, and comprised of 12 members. None of these panelists had previous experience in judging textural attributes of food, but all had previous experience in using the method of magnitude estimation.

Test samples. In order to determine if the effects observed in Experiment 2 were product-specific, a heterogeneous group of commodities was chosen for testing. Samples were chosen from the list of food items constituting the standard rating scales for mechanical texture attributes (Szczesniak et al., 1963). Some substitutions to these lists were made to compensate for current and local availability of items. The complete list of test items is shown in Table 2 for each standard scale.

Procedure. Six sessions were conducted in the same test booths as in Experiment 1. Each session was devoted to the testing of a single textural attribute. The six attributes were those of "hardness," "viscosity," "adhesiveness," "chewiness," "gumminess," and "fracturability." At the start of each session all panelists were given written instructions. In addition to containing a description of the task, the instructions contained an operational definition of the attribute to be judged during that session. The definitions for each attribute appear in Table 3. After reading the instructions and accompanying definition, each panelist was given a demonstration of the exact technique for judging the attribute. In addition, each panelist was observed while he/she made one or more trial evaluations using this technique.

After this orientation session, panelists were directed to the test booths to begin their evaluations. Panelists were allowed to keep the written instructions and definitions of the attributes with them in the booths and were encouraged to refer to them. The above procedures were all adopted to ensure, as best as possible, that consumers were defining and evaluating each textural attribute in the same manner as trained panelists.

All samples to be rated on the particular attribute were presented to panelists individually and in random order. After receiving the samples, panelists rated them using the method of modulus-free magnitude estimation. This scaling method was chosen because there were no differences between the two panels in their past

Table 1—Regression equations, Pearson product-moment correlation coefficients and coefficients of determination for the relationship between trained and consumer panel judgments of texture and appearance

	Regression equation[a]	Correlation coefficient	Coefficient of detm
Flakiness	T=2.42 C – 5.33	0.77**	0.59
Hardness	T=1.66 C – 3.44	0.75**	0.57
Chewiness	T=1.58 C – 2.14	0.84**	0.71
Fibrousness	T=1.52 C – 0.06	0.72**	0.52
Moisture	T=0.79 C – 0.62	0.53	0.28
Oily Mouthcoating	T=1.59 C – 0.87	0.75**	0.57
Darkness	T=1.04 C + 0.76	0.91**	0.82

[a] T = Trained Panel Ratings; C = Consumer Panel Ratings
** $p<0.01$

Table 2—Food items comprising the test samples used in Experiment 3

Product	Brand, type or preparation	Manufacturer	Sample size	Temp
Hardness				
Cream Cheese	Philadelphia	Kraft Foods	½" cube	45–55°F
Egg White	hard-cooked 5-min.		½" cube	room
Frankfurters	large, uncooked	Hebrew Nat'l	½" slice	50–65°F
Cheese	yellow, American, pasteurized process	Kraft Foods	½" cube	50–65°F
Olives	Spanish, stuffed	Durkee Famous Foods	1 olive, cut placed back to back	room
Peanuts	cocktail type in vacuum tin	Planters Peanuts	1 nut	room
Carrots	uncooked, fresh		½" slice	room
Peanut Brittle	candy part	Kraft Foods		room
Hard Candy	Charms	Charm Co.	1 piece	room
Viscosity				
Water	distilled		½ tsp	45–55°F
Light cream	Sealtest	Sealtest Foods	½ tsp	45–55°F
Heavy cream	Sealtest	Sealtest Foods	½ tsp	45–55°F
Evaporated milk		Carnation Co.	½ tsp	45–55°F
Maple Syrup	Vermont Maid	R.J. Reynolds Foods	½ tsp	45–55°F
Chocolate Syrup		Hershey Chocolate Corp.	½ tsp	45–55°F
Mixture: 1½ cup condensed milk & 1 tbl heavy cream		Borden Foods Sealtest Foods	½ tsp	45–55°F
Condensed milk		Borden Foods	½ tsp	45–55°F
Adhesiveness				
Hydrogenated vegetable oil	Crisco	Proctor & Gamble Co.	½ tsp	45–55°F
American Cheese		Kraft Foods	½" cube	45–55°F
Cream Cheese	Philadelphia	Kraft Foods	½" cube	45–55°F
Marshmallow topping	Fluff	Durkee-Mower	½ tsp	45–55°F
Peanut Butter	Skippy, smooth	Best Foods	½ tsp	45–55°F
Chewiness				
Rye bread	fresh, center cut	Arnold's Baking Co.	½" cube	room
Frankfurter	large, uncooked skinless	Hebrew National	½" slice	50–70°F
Cherry Red candy	Switzer Licorice	Beatrice Foods Co.	1 piece	room
Black Crows candy		Mason Candy Co.	1 piece	room
Caramel candy		Kraft Co.	1 piece	room
Tootsie rolls	midget size	Sweets Co. of America	1 piece	room
Gumminess				
40% flour paste	Gold Medal	General Mills	1 tbs.	room
45% flour paste	Gold Medal	General Mills	1 tbs.	room
50% flour paste	Gold Medal	General Mills	1 tbs.	room
55% flour paste	Gold Medal	General Mills	1 tbs.	room
60% flour	Gold Medal	General Mills	1 tbs.	room

(continued)

Table 2—Continued

Product	Brand, type or preparation	Manufacturer	Sample size	Temp.
Fracturability				
Corn muffin	Finast	First Nat'l Stores	½" cube	room
Eggs Jumbos		Stella D'Oro Biscuit Co.	½" cube	room
Graham crackers	Nabisco	National Biscuit Co.	½" square	room
Melba Toast		Devonsheer Melba Corp.	½" square	room
Rye Crisp	Finn Crisp	Vaasa Mills Ltd	½" square	room
Ginger snaps	Nabisco	National Biscuit Co.	½" square	room
Peanut brittle	Candy part	Kraft Foods	½" square	room

experience with this scaling technique. Each sample was judged once, after which the panelist expectorated, rinsed with distilled water, and awaited the next sample. A 90-sec interstimulus interval was maintained.

Data analysis. The magnitude estimation data were normalized (Stevens, 1971), and geometric means for trained and consumer panelists were calculated for each test sample. The means obtained from the trained panel for each item were regressed against the consumer panel means for these items in the same manner as described in Experiment 2.

Results

Table 4 shows the obtained linear regression equations relating trained panel judgments to consumer panel judgments for each attribute, as well as the corresponding correlation coefficients and coefficients of determination for each. All attributes were significantly correlated between panels. The high coefficients of determination observed in this experiment, as contrasted with Experiment 2, is likely due to the much wider range of each textural attribute represented in the test stimuli of this experiment. The limited range of each textural attribute in the fish samples of Experiment 2, probably resulted in statistical restriction of range. As was found in Experiment 2, the slopes (b) of the equations in Table 4 are all greater than 1.0. Thus, for all tested attributes, a greater range of perceptual differences were noted in the stimulus series by the trained panel than by the consumer panel. The fact that members of both panels had equivalent experience with the method of magnitude estimation supports the conclusion that these results are not due to differences in their experience with the scale type. Furthermore, since judgments of each attribute were made on a heterogeneous series of food products, the results cannot be attributed to differences in experience with a particular food product. Lastly, since both consumers and trained panelists used the same definitions and techniques for evaluating these attributes, the differences cannot be due to differences in the sensory attributes being evaluated. Rather, the results appear to be due to differences between consumer and trained panelists in either their awareness of, or sensitivity to, textural attributes, resulting from prior experience in judging these attributes in food.

Table 3—Definitions of textural attributes used in Experiment 3

Hardness	The perceived force required to compress a substance between the molar teeth.
Chewiness	The total perceived work required to masticate a sample to reduce it to a consistency suitable for swallowing.
Viscosity	The perceived force required to draw a liquid from a spoon over the tongue.
Gumminess	The perceived denseness that persists throughout mastication; the perceived energy required to reduce a semi-solid food to a state ready for swallowing.
Adhesiveness	The perceived force required to remove material that adheres to the mouth (generally the palate) during normal eating.
Fracturability	The perceived force with which a sample crumbles, cracks or shatters when a constant vertical force is applied to it.

Table 4—Regression equations, Pearson product-moment correlation coefficients and coefficients of determination for the relationship between trained and consumer panel judgements of each texture attribute

	Regression equation	Correlation coefficient	Coefficient of detm
Hardness	T=2.45 C − 14.51	0.99**	0.99
Chewiness	T=2.45 C − 3.14	0.99**	0.99
Fracturability	T=1.07 C − 0.36	0.98**	0.97
Gumminess	T=1.43 C − 7.36	0.98**	0.97
Viscosity	T=1.32 C − 1.12	0.97**	0.95
Adhesiveness	T=2.20 C − 12.85	0.93**	0.87

[a] T = Trained panel judgment; C = Consumer panel judgment
** $p<0.01$

EXPERIMENT 4

THE FOURTH EXPERIMENT was designed to further investigate the relationships between trained and consumer panel judgments of texture by specifically comparing the growth of perceived magnitude of textural attributes as a function of objective measures of the test samples. These data were collected as part of a more comprehensive study of the texture of bread, which has been reported elsewhere (Moskowitz et al., 1979a, b). From these data it was also possible to assess the relationship between consumer and trained panel judgments of acceptability.

Method

Panels. The first panel was the same trained texture profile panel described in Experiment 3, consisting of seven members for this series of tests. The second panel was a 28-member consumer panel, similar to that described in Experiment 1 – 3, with most having had experience in using the method of magnitude estimation. However, to ensure that all consumer panelists understood this scaling method, an orientation session (see below) preceeded the test sessions.

—Continued on next page

Test samples. Twelve rye breads were made by varying the percent rye flour (12, 19.5, 27, 42%) and the percent sucrose (0, 3, 6%) in a complete factorial design. These variations produced concomitant variations in the wheat flour level. All the breads also contained 1.5% shortening, 1% table salt, 38.7% water, and 0.8% yeast. The breads, in the form of 18-oz loaves, were baked in a rotary oven at 425°F for 30 minutes and were stored frozen (0°C) until used. Immediately prior to testing, the breads were removed from storage and thawed to room temperature. Samples were then tested with an Instron Universal Testing Machine to provide measures of their rheological properties (modulus of elasticity, hysteresis loss, strain energy, and stress at 60% strain). Density and moisture content were also determined.

Procedure. All consumers were given a 30-min orientation session on the day before the test to refresh them with the method of magnitude estimation and to give them practice in using this method to rate non-food stimuli. During the test session, the consumers rated each of the 12 samples of bread on each of ten textural attributes defined by the expert panel: "denseness," "firmness," "cohesiveness," "ease of particle removal," "roughness," "moistness," "cohesiveness during chewing," "adhesiveness," and "graininess." The definitions for each attribute appear in Table 5. In addition, both panels rated the breads for overall liking. Magnitude estimates were collected for each product on all attributes from both panels normalized to reduce interpanelist variability, panelist-by-panelist (Moskowitz, 1977), then averaged.

Table 5—Definitions of textural attributes for bread used in Experiment 4. Attributes are grouped into three categories: those which are evaluated initially and which involve surface characteristics; those evaluated upon first bite, using the incisor teeth; and those evaluated during mastication with the molar teeth

Surface

Ease of particle removal – The perceived ease with which particles (grains) can be removed from the surface of the sample, when sliding the tongue over the cut surface.

Roughness – The perceived roughness of the surface of the sample, when sliding the tongue over the cut surface.

First Bite

Firmness – The perceived force required to bite through the sample.

Cohesiveness – The perceived degree to which the sample holds together upon biting.

Denseness – The perceived degree to which the particles of bread are packed closely together.

Mastication

Moistness – The perceived degree of moisture in the sample.

Graininess – The number of grainy particles perceived in the mouth during initial stages of chewing.

Chewiness – The perceived effort required to prepare the sample to a state ready for swallowing.

Cohesiveness during chewing – The degree to which the sample holds together as a single mass during chewing.

Adhesiveness – The perceived degree to which the sample sticks to the teeth during chewing.

Results

The multivariate relationships between ingredient levels and rheological measures, between ingredient levels and trained panel ratings of specific textural attributes, between rheological measures and trained panel ratings, and between consumer ratings of liking and both ingredient levels and rheological measures have been reported elsewhere (Moskowitz, et al., 1979). Of interest here are the relationships between trained and consumer panel judgments of specific textural attributes and of liking. Because rheological measures were made on all test samples, a direct assessment and comparison of the growth of perceived magnitude as a function of stimulus intensity was made. Similarly, the relationships between liking and the rheological measures for both panels were also examined.

Intercorrelation matrices. Table 6 shows the intercorrelation matrices among sensory attributes for the trained and consumer panels. The patterns of correlations are similar for both panels. In particular the surface characteristics of "ease of particle removal" and "roughness" are positively correlated with one another and negatively correlated with all other attributes. "Firmness," "cohesiveness," "denseness," and "adhesiveness" are all highly positively correlated, while "moistness" is only correlated with the destructive attributes of "chewiness" and "cohesiveness during chewing." Of the total of 45 correlations, 19 were significant for the trained panel and 24 were significant for the consumer panel, suggesting somewhat more independence of responding for trained panelists.

Psychophysical functions. Although many of the correlations between texture attributes were high for both panels, such a situation could co-exist with a large absolute difference between panels in intensity ratings or in their rate of growth. Figures 5–12 show trained and consumer panel judgments of firmness, cohesiveness, denseness, and adhesiveness as a function of the modulus of elasticity and density of the breads, plotted in full logarithmic coordinates. These particular sensory and instrumental measures were chosen because of their presumed relationships with one another. Sensory attributes involving only surface characteristics of the breads or involving evaluation of destructive changes in the breads during mastication were not deemed theoretically appropriate for this analysis.

The equations describing the relationships in Figures 5–12 are power functions of the form $S = KI^n$, where S is the perceived magnitude of the textural attribute, I is the physical magnitude of the rheological property, n is the exponent of the function and k is a constant of proportionality. The value of n is of special interest, because it is an index of the rate of growth of perceived magnitude as a

—Text continued on page 1195

Table 6—Pearson product-moment correlation coefficients among all pairs of sensory texture attributes. The coefficients on the left of each cell are for the trained panel data. The coefficients on the right of each cell are for the consumer data

	Ease of particle removal	Roughness	Firmness	Cohesiveness	Denseness	Adhesiveness	Moistness	Chewiness	Cohesiveness during chewing
Roughness	0.97*/ 0.71*								
Firmness	–0.95*/–0.84*	–0.89*/–0.51							
Cohesiveness	–0.92*/–0.87*	–0.91*/–0.61	0.91*/ 0.95*						
Denseness	–0.96*/–0.88*	–0.94*/–0.50	0.95*/ 0.99*	0.92*/ 0.96*					
Adhesiveness	–0.73*/–0.77*	–0.72*/–0.54	0.72*/ 0.90*	0.74*/ 0.92*	0.84*/ 0.89*				
Moistness	–0.41 /–0.28	–0.46 /–0.54	0.35 / 0.40	0.30 / 0.47	0.51 / 0.36	0.57 / 0.66			
Chewiness	–0.68 /–0.82*	–0.60 /–0.64	0.69 / 0.92*	0.59 / 0.93*	0.78*/ 0.90*	0.71*/ 0.96*	0.77*/ 0.67		
Cohesiveness During Chewing	–0.51 /–0.66	–0.57 /–0.75*	0.30 / 0.71*	0.46 / 0.77*	0.57 / 0.66	0.66 / 0.79*	0.74*/ 0.74*	0.64 / 0.87*	
Graininess	0.14 / 0.66	0.55 / 0.65	–0.14 /–0.60	–0.20 /–0.60	–0.22 /–0.57	–0.22 /–0.64	–0.50 /–0.65	–0.44 /–0.77*	–0.45 /–0.79*

* $P<0.01$

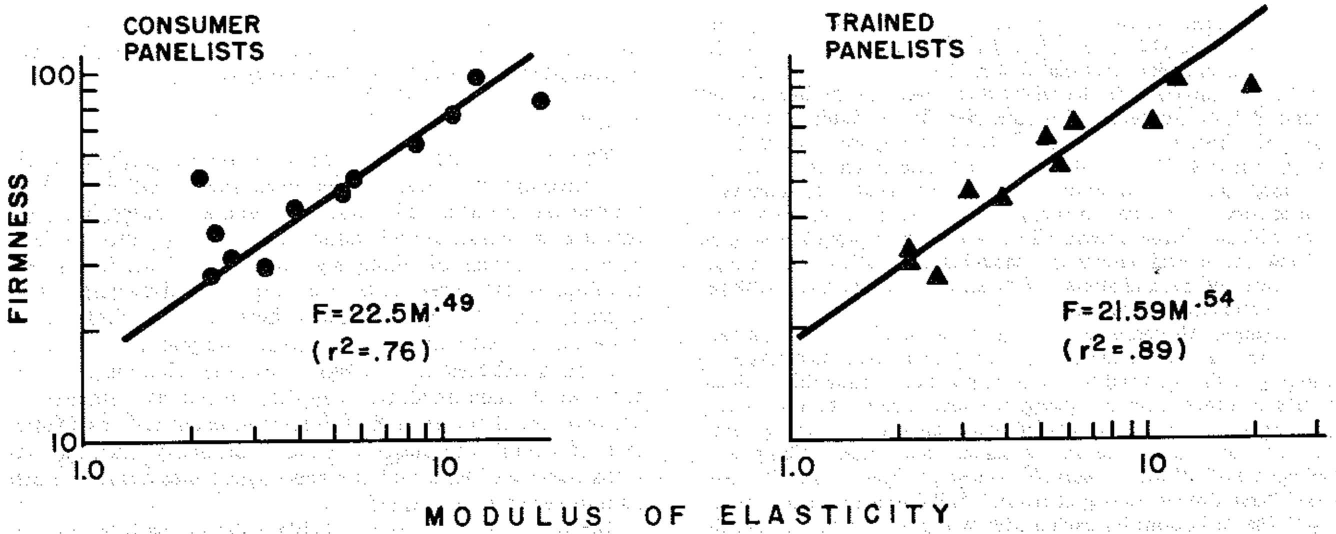

Fig. 5—Consumer (left) and trained (right) panel ratings of firmness as a function of the modulus of elasticity for the bread data of Experiment 4.

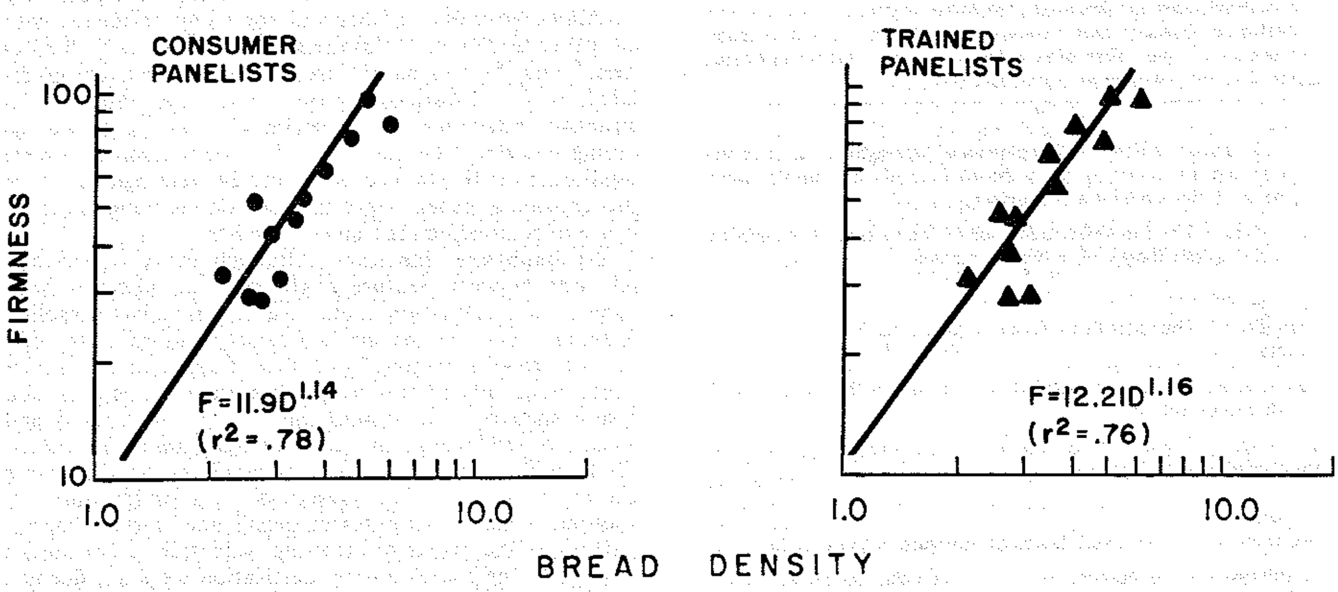

Fig. 6—Consumer (left) and trained (right) panel ratings of firmness as a function of bread density for the bread data of Experiment 4.

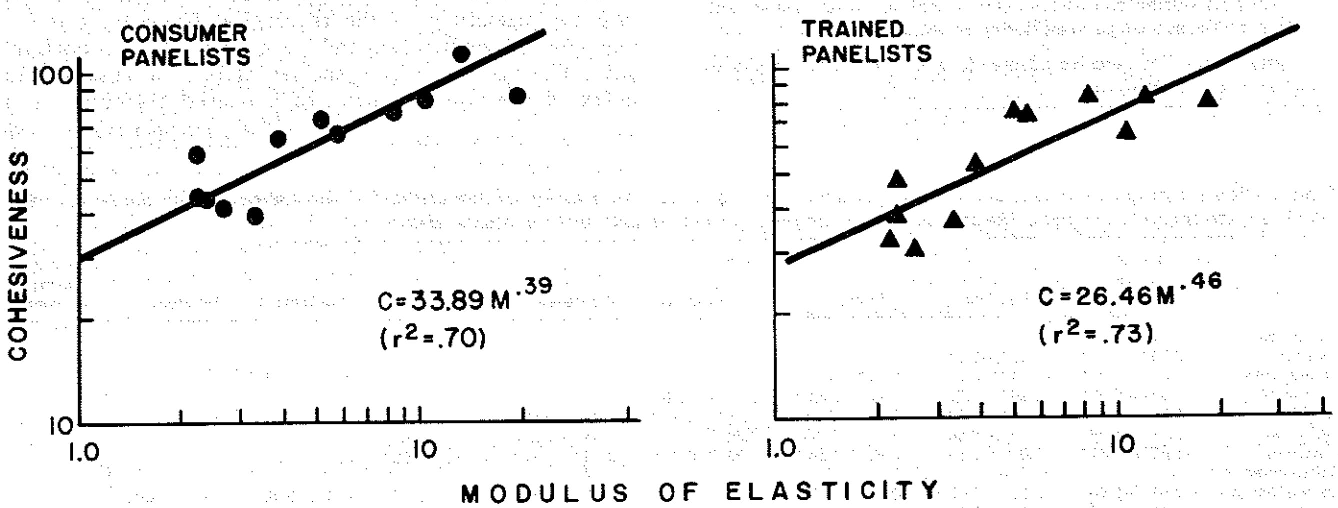

Fig. 7—Consumer (left) and trained (right) panel ratings of cohesiveness as a function of the modulus of elasticity for the bread data of Experiment 4.

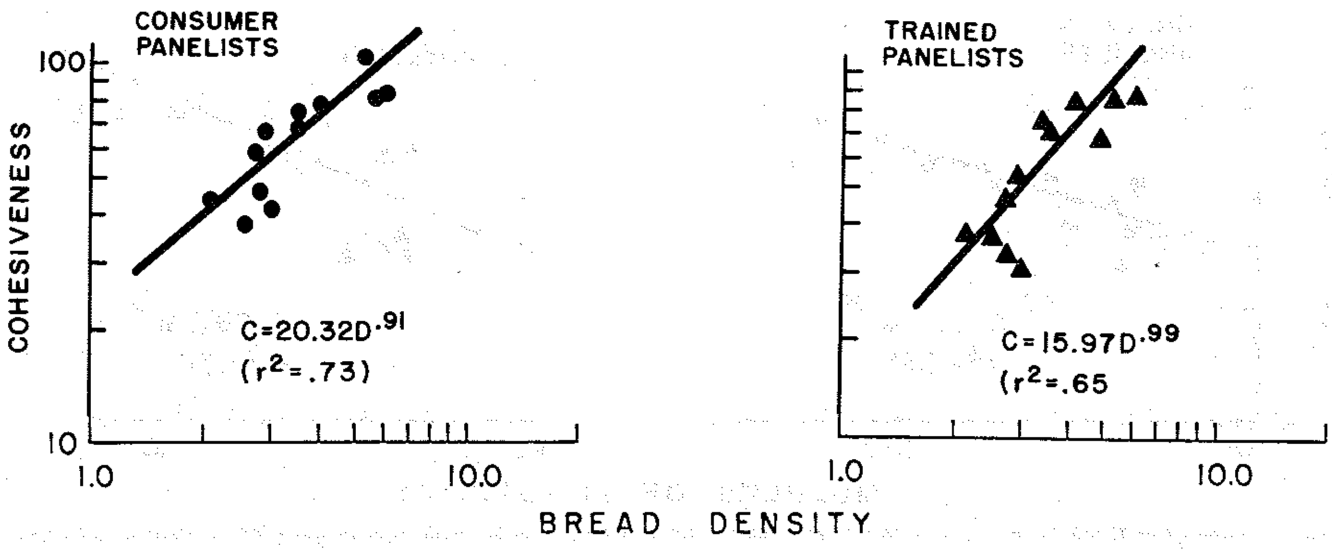

Fig. 8—Consumer (left) and trained (right) panel ratings of cohesiveness as a function of bread density for the bread data of Experiment 4.

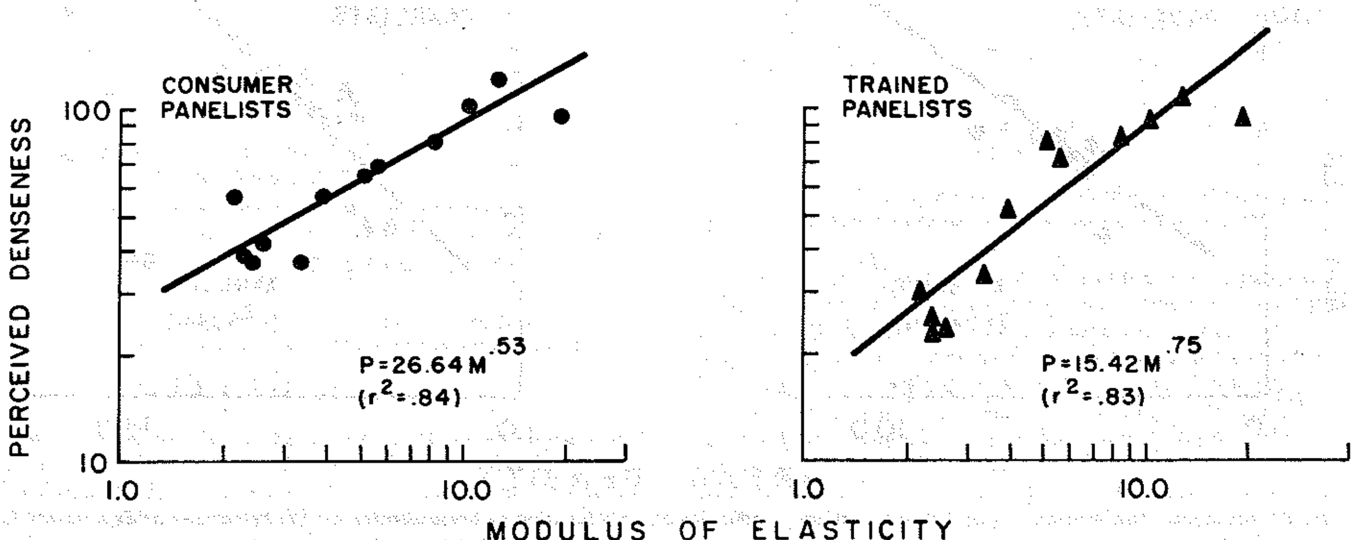

Fig. 9—Consumer (left) and trained (right) paned ratings of denseness as a function of the modulus of elasticity for the bread data of Experiment 4.

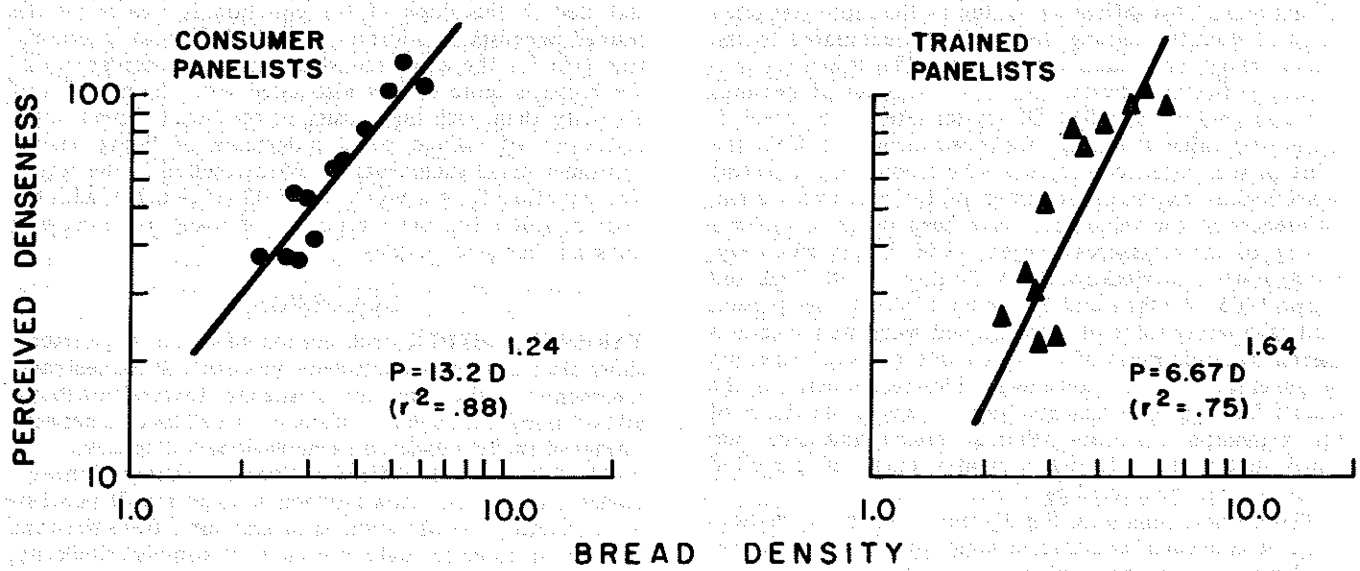

Fig. 10—Consumer (left) and trained (right) panel ratings of denseness as a function of bread density for the bread data of Experiment 4.

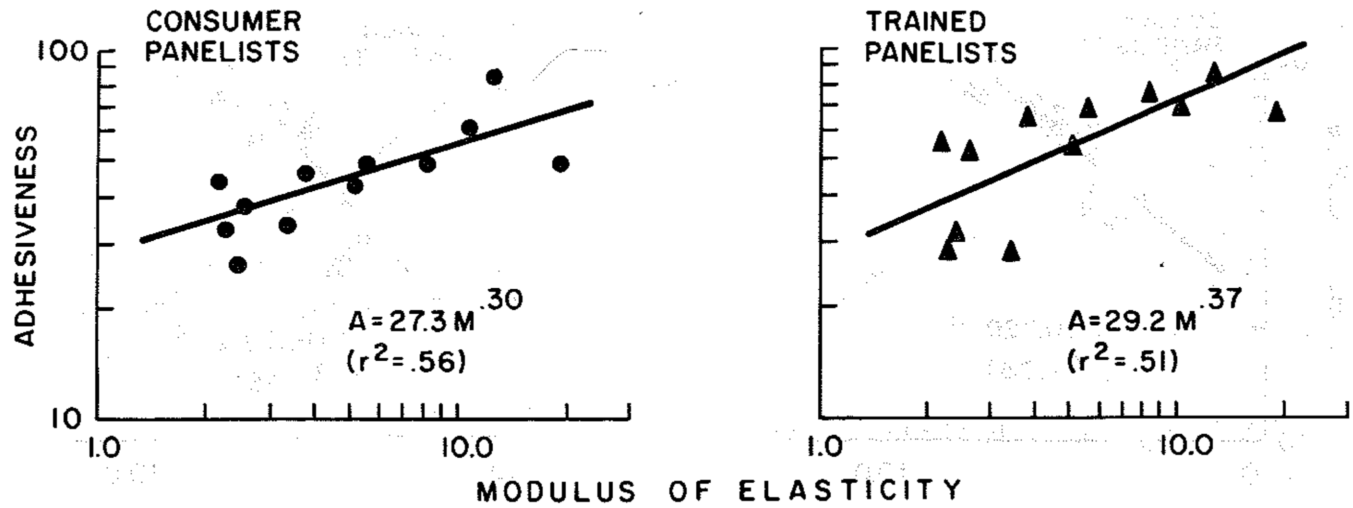

Fig. 11—Consumer (left) and trained (right) panel ratings of adhesiveness as a function of the modulus of ealsticity for the bread data of Experiment 4.

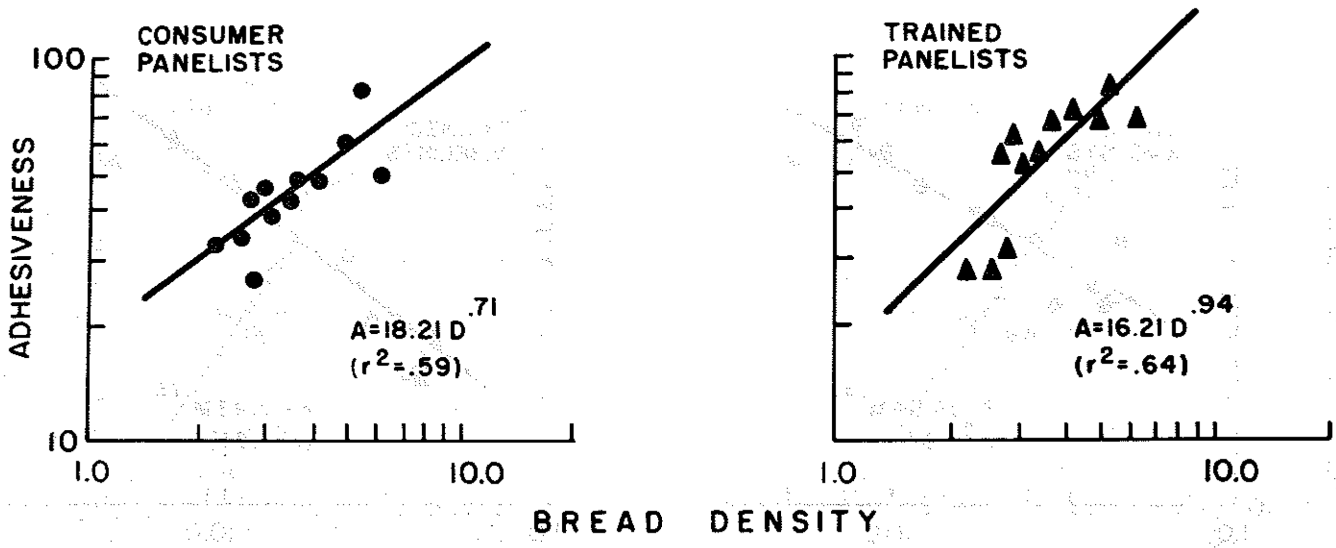

Fig. 12—Consumer (left) and trained (right) panel ratings of adhesiveness as a function of bread density for the bread data of Experiment 4.

function of increasing physical magnitude. In all cases, n is greater when the perceived magnitude is regressed against bread density than when it is regressed against the modulus of elasticity. This difference is due to the more restricted range of densities among the breads, as compared to the greater range of moduli of elasticity. The Pearson correlation coefficient between these two instrumental measures was high ($r=0.97$, $p<0.01$). Of greater interest, however, is the greater value of n noted for trained panelists. While the value of n is often considered to be a constant, reflecting physiological response properties of the receptor system, differences in the value of n have been reported under a variety of circumstances (Engen, 1956; Engen and Levy, 1958; Jones and Woskow, 1966; Poulton, 1968, Beck and Shaw; 1965; Pradhan and Hoffman, 1963). In the present study the larger value of n for trained panelists indicates a greater dynamic range of perceived intensities in this group of panelists than for consumers. This result parallels the results found in Experiments 2 and 3, where the slopes of the regression equations relating trained and consumer panel judgments reflected a greater range of perceived intensities for the trained panelists.

Pleasantness functions. Fig. 13 and 14 show the relationships of consumer and trained panel judgments of liking/disliking with the same objective measures above. Linear regression equations relating these measures show a decline in acceptability with increasing modulus of elasticity and/or density for both groups. However, the rate of decline, as reflected in the slope of the equation is greater for the trained panelists. In addition, the coefficients of determination (r^2) for the equations show that the acceptability of the bread is more closely associated with the modulus of elasticity than with the density of the bread. Direct linear regression of trained panel judgments of liking against consumer panel judgments of liking resulted in the regression equation $T_L = 1.66\ C_L - 35.08$ ($r^2 = 0.74$), with the slope of 1.66 reflecting a faster rate of change of liking with physical change in samples.

DISCUSSION

TAKEN TOGETHER, the results of these experiments show clear differences between consumer and trained panel judgments of the intensity of specific textural attributes and of liking, although no differences were observed between ratings of the two panels on a similarities scaling task.

The failure to find differences between ratings of the two panels on the similarities task was unexpected. If panelists are extracting specific features or attributes from the total percept in order to make estimates of stimulus similarity, it would be expected that the trained panelists would place

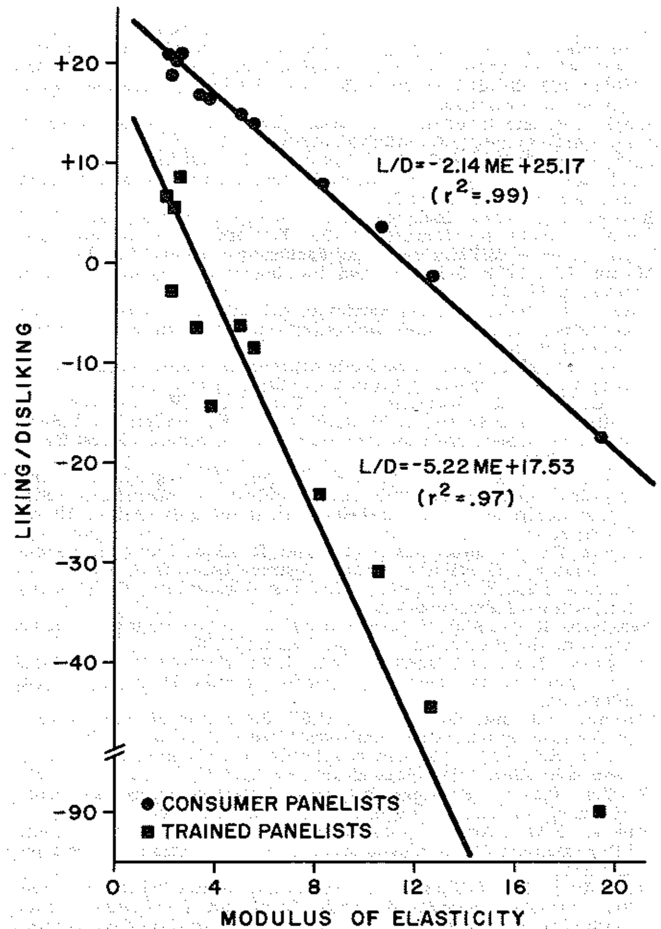

Fig. 13—Consumer and trained panel ratings of liking/disliking as a function of the modulus of elasticity for the bread data of Experiment 4.

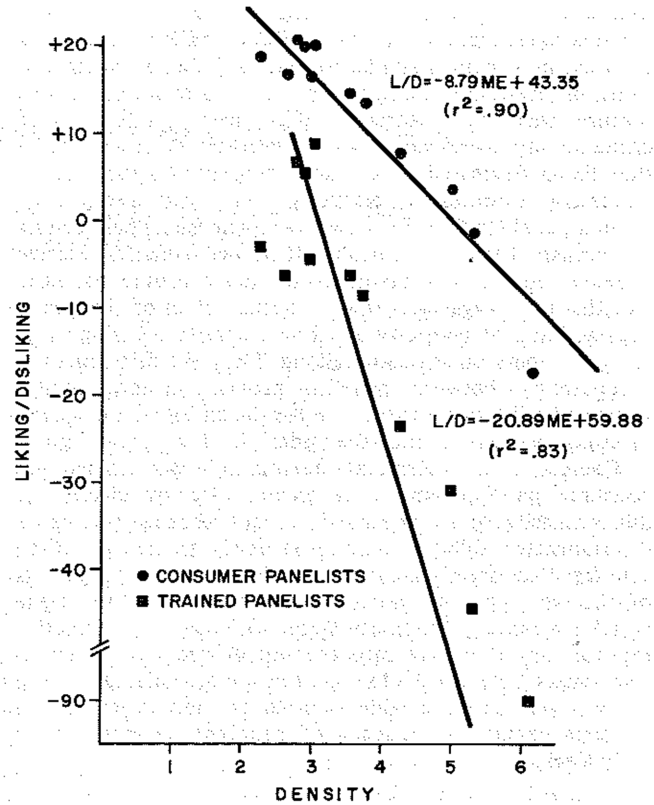

Fig. 14—Consumer and trained panel ratings of liking/disliking as a function of bread density for the bread data of Experiment 4.

more weight on the textural dimension of flakiness than consumers. This was not the case. However, the data from Experiments 2, 3 and 4 are clear in showing that the range of perceptual magnitudes for a variety of textural attributes is augmented in panelists who have undergone texture profile training. One possible explanation of the failure to observe such differences in the similarities scaling task is that the panelists are not operating in an analytic or "feature extracting" mode when judging similarities. That is, they are not cognitively aware of, and judging, each attribute independently. Thus, the major advantages of analytic training—commonality of attribute definition and commonality in procedure of evaluation of a single attribute—is lost. In the case of flakiness of fish, for example, judgment of this attribute by a trained panelist, will be guided by a specific pre-established definition of flakiness, such as "the perceived force required to produce separation between adjoining flakes within the muscle", and a specific method of evaluation, such as, "by manipulation of the muscle between tongue and palate." However, in a similarities task, the panelist neither considers each textural dimension of the sample sequentially nor considers the specific definition or method of evaluation for each. As a result, the analytic aspect is totally absent, and differences between trained panelists and consumers are not observed. Whether attribute definitions and methods of evaluation become more ingrained in trained panelists after many years of experience, so that training effects might appear in similarity judgments after ten or 15 yr of continued practice is a possibility that deserves further examination with a long-established panel.

The difference between consumer and trained panel judgments of specific textural dimensions in these experiments appear to be the result of an increased perceptual range for textural attributes resulting from training and experience. This notion is supported by the regression analyses of Experiment 2 and 3, in which the slope of the regression lines relating trained to consumer panel judgments were greater than 1.0 for all judged attributes. This notion was further supported by the results of Experiment 4, in which it was found that the exponents of the psychophysical functions relating perceived magnitude to underlying rheological measures were greater for trained panelists on all examined textural attributes.

These results parallel the results of numerous studies conducted over the years in a wide variety of senosry modalities, which have examined the effects of experience on perception. These studies, many of which have been catalogued and discussed in the text by Gibson (1969), have demonstrated an increased ability to discriminate among sensory stimuli following repeated experience with a specific stimulus or attribute domain. Examples of such perceptual learning in the food senses range from the ability of experienced wine-tasters to differentiate wines made from grapes of different vintage to the demonstration that taste thresholds for compounds characteristic of the four basic taste qualities can be significantly reduced by training (Pangborn, 1959). In the case of the present experiments, the experience that the trained panel had received improved their ability to discriminate among stimuli varying along given textural dimensions. This improved discriminative ability was manifested in an overall greater perceptual range of intensities for the stimulus series examined in these experiments. In addition, the present studies demonstrate the effect that such perceptual learning can have on judgments of liking for a product.

The observed difference between consumer and trained panel judgments of liking for bread in Experiment 4 was somewhat unexpected, since scaling of liking involves simultaneous assessment of all the sensory characteristics of a product, in a manner similar to that for judging similarities. Thus, differential training on one set of attributes, such as texture, was not expected to affect overall judgments of liking to any significant degree. Nevertheless, it was found that liking decreased much more quickly as a function of increasing modulus of elasticity and bread density for trained panelists than for consumer panelists. These results are consistent with the notion that trained panelists perceive a greater range of textural intensities than consumer panelists and that this exaggerates the perceptual effect of changes in the modulus of elasticity and bread density, causing exaggerated effects on associated liking. Thus, the differences in acceptability between panelists provide an independent source of validating information for the differences observed in response magnitudes of the underlying texture attributes.

Overall, these experiments demonstrate that trained and consumer panel judgments of texture differ quantitatively, due primarily to an expanded perceptual range (increased discriminative ability) on the part of the trained panelists. The fact that linear regression equations between judgments of the two types of panels account for up to 99% of the variability among judgments (Table 4) means that predictive relationships that take into account differences in perceptual ranges between the two groups, are possible. Data need to be obtained on a wider sample of food products and textures before the efficacy of panel-interchange can be fully assessed.

REFERENCES

Beck, J. and Shaw, W.A. 1965. Magnitude of the standard, numerical values of the standard, and stimulus spacing in the estimation of loudness. Percept. & Motor Skills 21: 151.

Brandt, M.A., Skinner, E.J., and Coleman, J.A. 1963. Texture profile method. J. Food Sci. 28: 404.

Bressan, L.P. and Behling, R.W. 1977. The selection and training of judges for discrimination testing. Food Technol. 31: 62.

Calvin, L.D. and Sather, L.A. 1959. A comparison of student preference panels with a household consumer panel. Food Technol. 13: 469.

Caul, J.F. 1957. The profile method of flavor analysis. Adv. in Food Res. 7: 1.

Civille, G.V. and Szczesniak, A.S. 1973. Guidelines to training a texture profile panel. J. Texture Studies 4: 204.

Cross, H.R., Moen, R., and Stanfield, M.S. 1978. Training and testing of judges for sensory analysis of meat quality. Food Technol. 32: 48.

Ellis, B.H. 1963. Chicanery in the flavor room. Proc. Amer. Soc. Brewing Chemists 37.

Engen, T. 1956. An evaluation of a method for developing ratio-scales. Am. J. Psychology 69: 92.

Engen, T. and Levy N. 1958. The influence of context on constant-sum loudness judgments. Am. J. Psychology 71: 731.

Gibson, E.J. 1969. "Principles of perceptual Learning and Development." Appleton-CenturyCrofts, New York.

Girardot, N.F., Peryam, D.R., and Shapiro, R. 1952. Selection of sensory testing panels. Food Technol. 6: 140.

Gruber, A. and Lindberg, B. 1966. Sensitivity, reliability and consumer taste testing. J. Marketing Res. 3: 235.

Hall, B.A., Tarver, M.G., and McDonald, J.G. 1959. A method for screening flavor panel members and its application to a two sample difference test. Food Technol. 8(12): 699.

Jones, F.N. and Woskow, M.J. 1966. Some effects of context on the slope in magnitude estimation., J. Exp. Psychology 71: 170.

Kapsalis, J. and Maller, O. 1981. Sensory and instrumental edibility measures for the grouping of fish species. Final Report to the Dept. of Commerce, National Marine Fisheries on Reimbursable Order #01-8-M01-6320, U.S. Army Natick Research & Development Laboratories.

Kiehl, E.R. and Rhodes, V.J. 1956. New techniques in consumer preference research. J. Farm. Econ. 38: 1335.

Kirkpatrick, M.E., Lamb, J.C., Dawson, E.H., and Eisen, J.N. 1957. Selection of a taste panel for evaluating the quality of processed milk. Food Technol. 11: 3.

Kluter, R.A. 1974. Relationships between laboratory and field testing. Activities Report of the Research and Development Associates for Military Food and Packaging Systems 26: 200.

Martin, S.L. 1973. Selection and training of sensory judges. Food Technol. 27: 22.

Miller, P.G., Nair, J.H., and Harriman, A.J. 1955. A household and a laboratory type of panel for testing consumer preference. Food Technol. 9: 445.

Moskowitz, H.R. 1977. Magnitude estimation. Notes on what, how, when and why we use it. J. Food Quality 1: 195.

Moskowitz, H.R., Kapsalis, J.G., Cardello, A.V., Fishken, D., Maller, O., and Segars, R. 1979a. Determining relationships among objective, expert, and consumer measures of texture. Food Technol. 33: 84.

Moskowitz, H.R., Cardello, A.V., Maller, O., Segars, R.A., and Kapsalis, J.G. 1979b. Product optimization: maximizing consumer acceptance and increasing profit margins. Bakers Digest 53(5): 8.

Nichols, T.L., Swanson, J.B., and Kluter, R. 1972. Group differences in sensory evaluation of food. Abstract listing No. 109, 1972 Annual Report, Pioneering Research Laboratory, U.S. Army Natick R&D Laboratories.

Pangborn, R.M. 1959. Influence of hunger on sweetness preferences and taste thresholds. Am. J. Clin. Nutrit. 7: 280.

Pangborn, R.M. and Dunkley, W.L. 1964. Laboratory procedures for evaluating the sensory properties of milk. Dairy Sci. Abstr. 26(2): 55.

Peryam, D.R. and Haynes, J.G. 1957. Prediction of soldiers' food preferences by laboratory methods. J. Appl. Psychol. 41: 2.

Poulton, E.C. 1968. The new psychophysics: Six models of magnitude estimation, Psychological Bull. 69: 1.

Pradhan, P.L. and Hoffman, P.J. 1963. Effect of spacing and range of stimuli on magnitude estimation. J. Exp. Psychol. 66: 533.

Sawyer, F.M., Stone, H., Abplanalp. H., and Stewart, G.F. 1962. "Repeatability" estimates in sensory panel selection. J. Food Sci. 27: 386.

Schlosberg, H., Pfaffmann, C., Cornsweet, J., and Pierrel, R. 1954. Selection and training of panels. In "Food Acceptance Testing Methodology — A Symposium," p. 45. Advisory Board on Quartermaster Research and Development Committee on Foods, National Academy of Sciences — National Research Council.

Simone, M., Leonard, S., Hinreiner, E., and Valdes, R.M. 1956. Consumer studies on sweetness of canned cling peaches. Food Technol. 10: 278.

Stevens, S.S. 1971. Issues in psychophysical measurement. Psych. Rev., 78: 426.

Stone, H., Sidel, J.L., Oliver, S., Woolsey, A., and Singleton, R.C. 1974. Sensory evaluation by Quantitative Descriptive Analysis. Food Technol. 28(11): 24.

Swartz, M.L. and Furia, T.E. 1977. Special sensory panels for screening new synthetic sweetness. Food Technol. 31: 51.

Szczesniak, A.S., Brandt, M.A., and Friedman, H.H. 1963. Development of standard rating scales for mechanical parameters of texture and correlation between the objective and the sensory methods of texture evaluation. J. Food Sci. 28: 397.

Szczesniak, A.S., Loew, B.J., and Skinner, E.Z. 1975. Consumer texture profile technique. J. Food Sci. 40: 1253.

Szczesniak, A.S. and Skinner, E.Z. 1973. Meaning of texture words to the consumer. J. Texture Studies 4: 378.

Wittes, J. and Turk, A. 1968. The selection of judges for odor discrimination panels. In "Correlation of Subjective-Objective Methods in the Study of Odors and Taste." ASTM STP 440, p. 49, Am. Soc. for Testing and Materials, Philadelphia.

Young, F.W. and Lewyckyj, R. 1979. ALSCAL—4 User's Guide. Carrboro, N.C.: Data Analysis and Theory Associates.

Zook, K. and Wessman, C. 1977. The selection and use of judges for descriptive panels. Food Technol. 31: 56.

Ms received 7/13/81; revised 1/13/82; accepted 1/14/82.